

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 5月24日

出 願 番 号

Application Number:

特願2000-153087

出 願 人

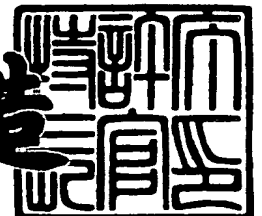
Applicant(s):

インターナショナル・ビジネス・マシーンズ・コーポレーション

2000年 9月18日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3074778

【書類名】 特許願

【整理番号】 JP9000042

【提出日】 平成12年 5月24日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/30

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 森本 康彦

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 手塚 集

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 日高 一義

【特許出願人】

 【識別番号】 390009531

 【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

 【識別番号】 100086243

 【弁理士】

 【氏名又は名称】 坂口 博

【復代理人】

 【識別番号】 100104880

 【弁理士】

 【氏名又は名称】 古部 次郎

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【選任した復代理人】

【識別番号】 100100077

【弁理士】

【氏名又は名称】 大場 充

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 領域算出方法、空間データマイニング装置、地図情報表示装置、空間データマイニングシステム、および記憶媒体

【特許請求の範囲】

【請求項 1】 住所などの空間情報を含むデータベースの中から 2 次元相関ルールを導き出して地図上で適用する領域算出方法であって、

前記 2 次元相関ルールを導き出すために調べる目的関数を定義するステップと

前記地図上の領域を所定の大きさからなるピクセルグリッドに分割するステップと、

分割された前記ピクセルグリッド毎に前記データベースからデータを集計するステップと、

定義された前記目的関数に基づいて前記ピクセルグリッドの領域を算出するステップと、

算出された前記領域に対応する前記地図上にあるエンティティを抽出するステップと、

抽出された前記エンティティに基づいて前記地図上で適用する領域を出力するステップとを含むことを特徴とする領域算出方法。

【請求項 2】 定義される前記目的関数は、出力の際に要求される領域情報が含まれていないことを特徴とする請求項 1 記載の領域算出方法。

【請求項 3】 前記ピクセルグリッドに分割するステップは、算出のための精度と計算時間とに基づいて前記所定の大きさを決定することを特徴とする請求項 1 記載の領域算出方法。

【請求項 4】 前記目的関数を最適化するような前記ピクセルグリッドの領域を算出するステップは、定義される制約と前記目的関数とに基づいて前記領域を算出することを特徴とする請求項 1 記載の領域算出方法。

【請求項 5】 前記地図上にあるエンティティを抽出するステップは、算出された前記ピクセルグリッドの外枠との距離および当該外枠に対する角度に基づいて当該外枠に対応するエンティティを抽出することを特徴とする請求項 1 記載

の領域算出方法。

【請求項 6】 前記領域を出力するステップは、抽出された前記エンティティを接続した領域を前記地図上に表示することを特徴とする請求項 1 記載の領域算出方法。

【請求項 7】 住所などの空間情報を含むデータベースの中から 2 次元相関ルールを導き出す領域算出方法であって、

前記 2 次元相関ルールを導き出すために調べる目的関数を定義するステップと、

前記データベースからデータを入力するステップと、

入力された前記データを集計して、定義された前記目的関数に適合する領域を算出するステップと、

地図情報を入力するステップと、

入力された前記地図情報から地図上にあるエンティティを抽出するステップと、

算出された前記領域に対応する領域エッジ候補を抽出された前記エンティティから選択するステップと、

選択された前記領域エッジ候補を入力された前記地図情報に反映して表示するステップとを含むことを特徴とする領域算出方法。

【請求項 8】 抽出される前記地図上にあるエンティティは、道路情報であることを特徴とする請求項 7 記載の領域算出方法。

【請求項 9】 前記領域エッジ候補を選択するステップは、前記エンティティを切断して当該領域エッジ候補を選択し、

当該領域エッジ候補を前記地図上に反映して表示するステップは、当該領域エッジ候補を連結して当該地図上に表示することを特徴とする請求項 7 記載の領域算出方法。

【請求項 10】 前記領域エッジ候補を前記地図上に反映して表示するステップは、当該領域エッジ候補を強調して当該地図上に表示することを特徴とする請求項 7 記載の領域算出方法。

【請求項 11】 住所などの空間情報を含むデータベースの中から 2 次元相

関ルールを導き出す空間データマイニング装置であって、

前記2次元相関ルールを導き出すために用いる目的関数を定義する定義手段と

地図情報を入力する入力手段と、

前記入力手段により入力された前記地図情報に基づいて地図上の領域を所定の大きさからなるピクセルグリッドに分割する分割手段と、

前記データベースからデータを入力すると共に、前記分割手段により分割された前記ピクセルグリッド毎に当該データを集計するデータ集計手段と、

前記データ集計手段による集計に基づいて、前記定義手段により定義された前記目的関数に適合する前記ピクセルグリッドの領域を算出するグリッド領域算出手段と、

前記入力手段により入力された前記地図情報から前記地図上にあるエンティティを抽出するエンティティ抽出手段と、

前記エンティティ抽出手段により抽出された前記エンティティの中から前記グリッド領域算出手段により算出された前記領域に対応するエッジ候補を選択するエッジ候補選択手段と、

前記エッジ候補選択手段により選択された前記エッジ候補に基づいて前記地図情報に対応した領域を出力する出力手段とを備えたことを特徴とする空間データマイニング装置。

【請求項12】 前記エッジ候補選択手段は、前記グリッド領域算出手段により算出された前記領域の外枠から所定距離内にあるエンティティを前記エッジ候補として選択することを特徴とする請求項1記載の空間データマイニング装置。

【請求項13】 前記エッジ候補選択手段により選択された前記エッジ候補を接続する接続手段とを更に備えたことを特徴とする請求項1記載の空間データマイニング装置。

【請求項14】 住所などの空間情報と関連付けされた属性情報を含むデータが格納されたデータベースからデータを入力するデータ入力手段と、

道路情報などの地図上のエンティティを含む地図情報を入力する地図情報入力

手段と、

分析が要求される目的関数を定義する目的関数定義手段と、

前記目的関数定義手段により定義された前記目的関数に基づいて前記データ入力手段により入力された前記データを集計する集計手段と、

前記集計手段による集計に基づいて、前記地図情報入力手段により入力された前記エンティティの中から特定のエンティティを選択する選択手段と、

前記地図情報から地図を表示すると共に、前記選択手段により選択された前記エンティティを強調して当該地図に表示する表示手段とを備えたことを特徴とする地図情報表示装置。

【請求項 1 5】 前記表示手段は、前記選択手段により選択された前記エンティティの中から所定の線分からなる有効エッジ候補を複数、選択すると共に、選択された複数の当該有効エッジ候補を連結させて表示することを特徴とする請求項 1 4 記載の地図情報表示装置。

【請求項 1 6】 住所などの空間情報と関連付けされた属性情報を含むデータが格納されたデータベースと、

道路情報などの地図上のエンティティを含む地図情報が格納された地図情報メモリと、

分析を要求するための目的関数を入力する入力装置と、

前記データベースから前記データを入力すると共に、前記入力装置により入力された前記目的関数に基づいて当該データを集計する集計手段と、

前記地図情報メモリから、前記地図上に存在する所定のエンティティを入力すると共に、前記集計手段による集計に基づいて、当該エンティティの中から特定のエンティティを選択する選択手段と、

前記選択手段により選択された前記エンティティを調整する調整手段と、

前記調整手段により調整された前記エンティティの情報を前記地図情報メモリに格納された地図と共に表示する表示装置とを備えたことを特徴とする空間データマイニングシステム。

【請求項 1 7】 領域情報が与えられていない目的関数に基づいて住所などの空間情報を含むデータベースの中から所定の領域を導き出すプログラムを格納

した記憶媒体であって、

前記プログラムは、前記データベースからデータを入力するステップと、入力された前記データを集計して前記目的関数を最適化する領域を算出するステップと、地図情報を入力するステップと、入力された前記地図情報から地図上にあるエンティティを抽出するステップと、算出された前記領域に対応する領域候補を抽出された前記エンティティから選択するステップと、選択された前記領域候補を入力された前記地図情報に反映して出力するステップとを含むことを特徴とする記憶媒体。

【請求項 1 8】 前記プログラムにおける前記最適化する領域を算出するステップは、前記地図上を所定の大きさからなるピクセルグリッドに分割し、分割された当該ピクセルグリッド毎に前記データを集計することを特徴とする請求項 1 7 記載の記憶媒体。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、空間データマイニングにおけるデータベース処理に関し、より詳しくは、空間データマイニングの基本機能である最適領域を算出する手法、装置等に関する。

【 0 0 0 2 】

【従来の技術】

データベースにある住所などの空間情報を空間的な文脈で解釈し、その空間的な法則を大量のデータの中から導き出す新しい技術を空間データマイニングと呼ぶ。この空間データマイニングは、大量データに対して、コストの高い空間・幾何学演算を行なう必要があり、技術的にはかなり難しい問題に直面する場合が多い。そのために、空間データマイニングはあまり発展しているとは言えず、未開拓の研究分野である。しかしながら、その一方で、空間データマイニングは、情報産業の中でも最もビジネスボリュームの大きなデータベースである G I S (Geographical Information System: 地理情報システム) 分野を大いに発展させる可能性を秘めた基礎技術と考えられ、ビジネス的、また技術的にも大変有望な技術

分野である。

【 0 0 0 3 】

ここで、空間的な解析機能を持たないデータベースシステムでは、データベースの属性から分類できる集計操作だけしか行なうことができない。このような場合には、例えばデータベースに行政区域を示す情報があるときは市町村行政区域毎に、例えば、データベースに店舗を示す情報がある場合には、店舗毎の集計操作しか行なうことができない。

【 0 0 0 4 】

データを空間的に解析することは、例えば、エリアマーケティングなどでは必要不可欠であり、最先端のGISでは、空間情報を使用して解析した結果とデータベース演算の結果とを統合して計算する機能を実現するものが存在する。即ち、空間情報を用いた演算結果からデータベース上の注目する属性値を集計する操作は、例えば、主要道路で分けられるブロック毎の集計操作や店舗の勢力圏毎の集計操作等、データベースと統合化されたGISにて実現することが可能である。

尚、出願人は、先に特願平09-034605号(米国特許第5,991,752号)にて、データ間の結合ルールを人間がより把握しやすい形で提示し、多くの結合ルールを可視化することにより、使用する人間の選択の幅を増大させ、より重要な結合ルールを見いだすことを可能とする技術について提案している。

【 0 0 0 5 】

【発明が解決しようとする課題】

上述の例のように、最先端の統合型GISでは、空間的にデータの集計を行なって、その集計値の高いものを取り出すことで、ある程度、空間的な特徴や傾向を見るといったデータマイニング的な分析が可能となる。

しかしながら、将来的な空間データマイニングの分野では、更に高度な分析が要求されている。その1つが領域最適化である。例えば、高価な商品の営業戦略を立てる場合に「平均年収を最大化あるいは最小化する領域」を知っておくことは非常に重要であるし、有効な警備体制を整えたい場合に「犯罪発生率を最大化する領域」を把握したいという要望が強い。上述の従来技術を応用し、例えば計

算可能な領域毎に集計値を計算してその値の高いものを取り出すように構成することも可能であるが、これらの従来技術の応用範囲では、空間的な連続性を考慮することができない。その結果、得られた領域から上述の例のような何らかの空間的な法則を発見することができず、将来的な空間データマイニングのための機能としては不十分なものである。

【 0 0 0 6 】

本発明は、以上のような技術的課題を解決するためになされたものであって、その目的とするところは、空間的な連続性を考慮した上で領域最適化を図り、より高度な空間データマイニングを提供することにある。

また他の目的は、最適化された領域の算出を高速に行なうことにある。

更に他の目的は、実在の地図エンティティに合うように調整して、算出された領域情報を地図と共に出力することにある。

【 0 0 0 7 】

【課題を解決するための手段】

かかる目的のもと、本発明は、位置情報などの空間情報を含む大量のデータベースから、出力領域に対する制約(例えば、最小あるいは最大面積、最小あるいは最大人口など)と目的関数値(例えば、ある商品の売り上げ総数、犯罪の発生件数など)に基づいて2次元相関ルールを求め、地図上に適用させるものである。

即ち、本発明は、住所などの空間情報を含むデータベースの中から2次元相関ルールを導き出して地図上で適用する領域算出方法であって、この2次元相関ルールを導き出すために調べる、出力の際に要求される領域情報が含まれていない目的関数を定義するステップと、地図上の領域を所定の大きさからなるピクセルグリッドに分割するステップと、分割されたピクセルグリッド毎にデータベースからデータを集計するステップと、定義された目的関数に基づいてその目的関数を最適化するピクセルグリッドの最適連結領域等の領域を算出するステップと、算出された最適連結領域等の領域に対応する地図上にあるエンティティを抽出するステップと、抽出されたエンティティに基づいて地図上で適用する領域を出力するステップとを含むことを特徴としている。

【 0 0 0 8 】

ここで、地図上のエンティティとは、地図上に現われる道路や川、湖、公園、森、行政区画などの実体が該当する。

また、ピクセルグリッドに分割するステップは、算出のための精度と計算時間とに基づいて所定の大きさを決定することを特徴とすることができる。即ち、算出のための精度を高める場合は、ピクセルグリッドを細かく分割する方法が挙げられ、算出処理に要する時間を短くするためには、大まかなピクセルグリッドを採用することが好ましい。

更に、目的関数を最適化するようなピクセルグリッドの領域を算出するステップは、定義される制約と目的関数とに基づいて領域を算出することを特徴とすることができる。この制約とは、領域最適化のために、例えば、最小顧客数等の他、ピクセルグリッドに分割した場合にはそのピクセルグリッド数等が該当するであろう。

【 0 0 0 9 】

また、地図上にあるエンティティを抽出するステップは、算出されたピクセルグリッドの外枠との距離および外枠に対する角度に基づいて外枠に対応するエンティティを抽出することを特徴とすることができる。例えば、外枠との距離が近くても、その外枠との角度が大きく、外枠との関連性が低いと判断できるエンティティは除外されることが好ましい。

また更に、領域を出力するステップは、エンティティを接続した領域を地図上に表示することを特徴とすれば、ユーザに対して提供する形態として、好ましいものとなる。但し、必ずしも地図と共に出力せずに、例えば領域を文言にて表現して出力することも可能である。

【 0 0 1 0 】

一方、他の観点から発明を捉えると、本発明は、住所などの空間情報を含むデータベースの中から2次元相関ルールを導き出す領域算出方法であって、2次元相関ルールを導き出すために調べる目的関数を定義するステップと、住所情報等の空間情報やその他の属性を含むデータベースからデータを入力するステップと、入力されたデータを集計して、定義された目的関数を最適化するためにこの目的関数に適合する領域を算出するステップと、地図情報を入力するステップと、

入力された地図情報から地図上にある道路、河川、市町村境界線、海岸線などの地図情報等のエンティティを抽出するステップと、算出された領域に対応する領域エッジ候補を抽出されたエンティティから選択するステップと、選択された領域エッジ候補を入力された地図情報に反映して表示するステップとを含むことを特徴とすることができる。

【0011】

この領域エッジ候補を選択するステップは、エンティティを切断して領域エッジ候補を選択し、この領域エッジ候補を地図上に反映して表示するステップは、領域エッジ候補を連結して地図上に表示することを特徴とすることができる。更に、切断したエンティティに、「ひげ」（複数のエンティティが交差した先に伸びる行き止まりになる折れ線）が存在する場合に、その「ひげ」を解消して、無駄のない閉領域を形成するように構成すれば、ユーザに対して見易い、より美しい出力を提供できる点で優れている。

また、領域エッジ候補を地図上に反映して表示するステップは、領域エッジ候補に対してその外枠の色付けや太線、閉領域の色分け等によって強調して地図上に表示することを特徴とすれば、ユーザに対して更に見易いアウトプットを提供することができる点で好ましい。

【0012】

更に、上記目的を達成するために、本発明は、住所などの空間情報を含むデータベースの中から2次元相関ルールを導き出す空間データマイニング装置であって、2次元相関ルールを導き出すために用いる目的関数を定義する定義手段と、地図情報を入力する入力手段と、入力された地図情報に基づいて地図上の領域を所定の大きさからなるピクセルグリッドに分割する分割手段と、データベースからデータを入力すると共に、分割手段により分割されたピクセルグリッド毎にデータを集計するデータ集計手段と、この集計に基づいて、定義された目的関数に適合するピクセルグリッドの領域を算出するグリッド領域算出手段と、入力された地図情報から地図上にあるエンティティを抽出するエンティティ抽出手段と、抽出されたエンティティの中からグリッド領域算出手段により算出された領域に対応するエッジ候補を選択するエッジ候補選択手段と、選択されたエッジ候補に

基づいて地図情報に対応した領域を出力する出力手段とを備えたことを特徴としている。

【0013】

このエッジ候補選択手段は、グリッド領域算出手段により算出された領域の外枠から所定距離内にあるエンティティをエッジ候補として選択することを特徴とすることができる。この所定距離内とは、例えばグリッドエッジを囲むグリッドを一辺とするピクセルグリッドを含む領域とすることができる。

また、エッジ候補選択手段により選択されたエッジ候補を補間関数などを用いて接続する接続手段とを更に備えたことを特徴とすることができる。

【0014】

他の観点から捉えると、本発明が適用される地図情報表示装置は、住所などの空間情報と関連付けされた属性情報を含むデータが格納されたデータベースからデータを入力するデータ入力手段と、道路情報などの地図上のエンティティを含む地図情報を入力する地図情報入力手段と、分析が要求される目的関数を定義する目的関数定義手段と、定義された目的関数に基づいてデータ入力手段により入力されたデータを集計する集計手段と、この集計手段による集計に基づいて、地図情報入力手段により入力されたエンティティの中から特定のエンティティを選択する選択手段と、地図情報から地図を表示すると共に、選択手段により選択されたエンティティを強調して地図に表示する表示手段とを備えたことを特徴とすることができる。この表示手段は、エンティティの中から所定の線分からなる有効エッジ候補を複数、選択すると共に、選択された複数の有効エッジ候補を連結させて表示することを特徴とすることができる。

【0015】

一方、本発明が適用される空間データマイニングシステムは、住所などの空間情報と関連付けされた属性情報を含むデータが格納されたデータベースと、道路情報などの地図上のエンティティを含む地図情報が格納された地図情報メモリと、分析を要求するための目的関数を入力する入力装置と、データベースからデータを入力すると共に、入力された目的関数に基づいてデータを集計する集計手段と、地図情報メモリから、地図上に存在する所定のエンティティを入力すると共

に、集計手段による集計に基づいて、エンティティの中から特定のエンティティを選択する選択手段と、選択されたエンティティを調整する調整手段と、調整されたエンティティの情報を地図情報メモリに格納された地図と共に表示する表示装置とを備えたことを特徴としている。このデータベースや地図情報メモリは、例えば表示装置等とは離れた場所に格納されている場合の他、本方式を処理するコンピュータのハードディスクドライブやCD-ROM等に格納されている場合がある。

【 0 0 1 6 】

また、本発明は、領域情報が与えられていない目的関数に基づいて住所などの空間情報を含むデータベースの中から所定の領域を導き出すプログラムを格納した記憶媒体であって、このプログラムは、データベースからデータを入力するステップと、データを集計して目的関数を最適化する連結ピクセル領域を算出するステップと、地図情報を入力するステップと、地図情報から地図上にあるエンティティを抽出するステップと、算出された領域に対応する領域候補を抽出されたエンティティから選択するステップと、選択された領域候補を入力された地図情報に反映して出力するステップとを含むことを特徴としている。この記憶媒体としては、例えばCD-ROM等の可搬性のある媒体の他、ネット等を介してプログラムをダウンロードするための、プログラムを提供する側のハードディスク等の記憶媒体、ダウンロード等によってプログラムの提供を受けたユーザ側のハードディスク等の記憶媒体を含むものである。

ここで、このプログラムにおける最適化する領域を算出するステップは、地図上を所定の大きさからなるピクセルグリッドに分割し、分割されたピクセルグリッド毎にデータを集計することを特徴とすれば、算出精度を高めることができると共に、処理時間を短縮して出力することが可能となる。

【 0 0 1 7 】

【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいてこの発明を詳細に説明する。

図1は、本実施の形態における最適領域算出手法を実現するためのコンピュータ・システムの概略構成を説明するための図である。本実施の形態における処理

アルゴリズムは、図示するようなコンピュータ・システムにおいて実行可能なプログラムとすることができる。この処理プログラムは、ハードディスクドライブ(HDD)15に格納され、実行時にはメインメモリ12にロードされてCPU11によって処理される。また、このHDD15には大量のデータベースが格納されている。格納されるデータベースとしては、GIS画面の地図情報、地図内にある地図エンティティ(湖、海、森、道路、河川など)の情報であるスペシャルデータ、住所情報と共に例えば買上金額や年収情報等からなる顧客データ等がある。HDD15に格納される処理プログラムはそのデータベースに対するアクセスを行うものである。

【0018】

地理情報システム(GIS)画面等の地図情報や最適化領域の計算結果は、表示装置16によってユーザに提示される。ユーザは、入力装置17にて調べたい目的関数の入力や、データ出力の命令等を入力する。このような入力装置17には、キーボードやマウス、ポインティング・デバイスやディジタイザ等が含まれる。更に、出力結果を補助記憶装置であるフロッピーディスクドライブ(FDD)13のフロッピーディスクに記憶したり、また新たなデータをFDD13から入力することも可能である。また、CD-ROMドライブ14を用いてCD-ROMに格納された地図情報やデータを入力することもできる。

【0019】

更に、本実施の形態を実現したコンピュータ・プログラムは、フロッピーディスクやCD-ROMといった記憶媒体に記憶して持ち運ぶことができる。この場合、通常のデータベース検索プログラムのデータ取り出し部分や、表示装置16に表示するだけの処理を行うプログラムは、すでにHDD15に記憶されている場合もある。従って、それ以外の部分が上記のような各種記憶媒体にて流通することは通常行われる事項である。また、図示されていない通信装置がバス18に接続されて、遠隔地にあるデータベースを用いて処理したり、処理結果を遠隔地に送信するように構成することも可能である。即ち、住所などの空間情報を含む大量のデータベースや地図情報を図1に示す構成の外部に設けるように構成することもできる。

【0020】

図2は、本実施の形態における最適領域算出方法を実現するための処理の流れを示した図であり、主にCPU11における動作である。まず、HDD15等のメモリ媒体から地図データが読み込まれ、読み込まれた地図に対して等間隔あるいはデータ数間隔にてバケット化される(ステップ101)。次に、HDD15等のメモリ媒体から顧客データを読み込み、バケットへの振り分けを行ない(ステップ102)、振り分けられたバケット毎にデータの集計を行なう(ステップ103)。次に、目的のグリッド領域を求めるためのアルゴリズムからなる領域切り出しエンジンによって、例えばユーザによって入力装置17から入力された目的関数値(例えば、ある商品の売り上げ総数、犯罪の発生件数等)から得られる最適化関数を最も良くする領域が算出される(ステップ104)。この最適化関数としては、例えば、エントロピー関数、GINIインデックス、 χ (カイ)2乗、平均2乗誤差等がある。

【0021】

次に、隣接ピクセル領域の幅を指定する等、隣接ピクセル領域の計算がなされる(ステップ105)。その後、HDD15等のメモリ媒体から地図エンティティを読み出し、隣接ピクセルと交差する地図エンティティの切り出しを行なう(ステップ106)。次に、切り出された線分(以降、「候補線分」と呼ぶ)から「有効エッジ候補」を選択する(ステップ107)。その後、連続していない「有効エッジ候補」の線分を接続し(ステップ108)、クロスした線分(クロス線分)の先に伸びる行き止まりになる折れ線である「ひげ」を除去する(ステップ109)。以上のようにして得られた結果は、表示装置16に対して地図情報と共に表示出力することが可能である。

【0022】

図3(a),(b)は、HDD15等に格納されている顧客データベース例と入力装置17から入力される制約、目的関数の例を示す図である。図3(a)では、あるデパートの顧客データベース(DB)の一例を示しており、顧客IDや、名前、住所、売上金額、年収等の属性に基づいて顧客データが格納されている。また、図3(b)の例では、データベース上に定義されている空間情報と関係付けられた

属性などを用いて定義される領域に対する制約と目的関数が定義されている。

【 0 0 2 3 】

図 4 (a) , (b) は、図 2 に示したステップ 1 0 1 にて読み込まれる地図とバケット化された地図の例を示す図である。図 4 (a) に示す地図は、GIS の地図データとして HDD 1 5 等に格納されている。この地図を読み込み、この地図上で適当に距離を決め、図 4 (b) に示すように、その距離の長さの辺を持つ正方形を単位とするピクセルグリッドを計算範囲の対象となる地図上に作る。このピクセルグリッドの大きさは、どの位の精度で計算したいか、に依存し、計算時間とのトレードで決定されるものである。即ち、ピクセルグリッドを大きくすると計算時間は速くなるが算出精度は荒くなり、逆に、ピクセルグリッドを小さくすると計算時間は遅くなるが算出精度は高くなる。

【 0 0 2 4 】

図 5 は、図 2 に示したステップ 1 0 3 にて、ピクセルグリッド上のバケットに集計されたデータの一例を示すものである。図 2 に示したステップ 1 0 2 で、各入力レコードに、そのレコードの座標値に応じたバケットのインデックス値が割り当てられ、その結果に基づいて各バケット毎にデータが集計される。ここでは、データベースを 1 度スキャンして、上述した制約と目的関数とを計算するのに必要な情報を各ピクセル毎に集計している。例えば、図 3 (b) に示した 2 つの制約・目的関数の例では、制約としてのグリッド数(ピクセル数)、各ピクセル毎の顧客数や、目的関数としての売上金額総和、年収額総和が集計される。この集計結果は、例えば図 5 に示すように、バケット番号として i 行 j 列 $[i, j]$ の表示 $([0, 0], [0, 1] \cdots [2, 0] \cdots \text{等})$ 、該当する顧客数(入力レコード数)、年収額総和で表わされ、HDD 1 5 やメインメモリ 1 2 等に格納される。この集計を高速に行なうためには、各ピクセルの情報はタテ・ヨコともに木構造で管理することが好ましい。即ち、どのピクセルに存在するかを位置決めするために、バイナリツリーで管理すれば、最悪でも木の深さに比例する計算時間で集計を行なうことが可能である。

【 0 0 2 5 】

図 6 (a) , (b) は、最適化関数値を最も良くする領域の算出例を示す図である。

。即ち、図2のステップ104に示した領域切り出しエンジンの一例を説明するものである。ステップ103で集計操作の終わった地図上のピクセルグリッド上で作られる任意の連続グリッド領域は、各ピクセルに集計した値の総和たる2次元平面上の点として特徴付けることができる。例えば、「領域の顧客数、領域の年収額総和」、「グリッド数、領域の売上金額総和」等で任意の領域が特徴付けられる。この領域切り出しエンジンについては、出願人による先の出願である特願平09-034605号に詳しいが、ここではその概略について説明する。

【0026】

図6(a)では、上述の例である「領域の顧客数、領域の年収額総和」の特徴を、横軸に領域の顧客数を、縦軸に領域の平均年収を取って表現している。ここで示される最小顧客数は、前述した領域最適化のための制約である。また、図6(b)では、「グリッド数、領域の売上金額総和」の特徴を、横軸にグリッド数を、縦軸に領域の売上金額総和を取って表現している。ここで示される最大グリッド数は、領域最適化のための制約である。全ての領域は、このような「領域の顧客数、領域の年収額総和」などの各ピクセルに集計されている2つの属性値で定義される平面の点集合である。この点集合の凸包上に存在する点については、かなり高速に計算することができる。即ち、凸包上の点は、ある傾き θ を有する接線と凸包との交点と考えることができる。図6(a)では、制約である最小顧客数の線より右で傾き最大の直線との接点で算出され、図6(b)では、制約である最大グリッド数の線より左で傾き最小の直線との接点で算出される。ここで、横にN、縦にNである N^2 個のピクセルグリッド上で考えると、そのような交点(接点)は、X単調領域で $O(N^2)$ 、直交凸領域で $O(N^3)$ の時間で計算できる。このように、ある傾き θ を有する接線と凸包との交点を求める手法を繰り返して適用することで、目的である最適に近いグリッド領域を高速に求めることができる。複数の領域を求めたい場合で、2番目、3番目以降の領域を求める場合は、既に求めた領域に含まれる各ピクセルのY値を目的関数でネガティブに作用するような値に書き換えて、同様な操作で求めることが可能である。

【0027】

図7は、算出して求められた最適ピクセルグリッド領域を示す図である。上述

のような作業によって図 7 に示すような太線の領域が抽出され、この太線の領域が最適化関数を最も良くするピクセルグリッド領域 3 1 となる。

図 8 は、図 2 のステップ 1 0 5 における隣接ピクセル領域の計算結果を示す図である。ここでは、ピクセルグリッド領域 3 1 の外枠であるエッジを挟む領域が計算され、図の斜線で示される隣接ピクセル領域 3 2 が求められる。このとき、外枠であるエッジを何個のピクセルで挟むかは、ユーザが入力装置 1 7 等から指定することも可能である。

【 0 0 2 8 】

図 9 (a) , (b) は、図 2 のステップ 1 0 6 における隣接ピクセルと交差する地図エンティティの切り出し作業を説明するための図である。図 9 (a) に示すように、地図内にあるポリゴン、折れ線で表示される地図エンティティ (湖、海、森、河川、道路、行政区画など) でピクセルグリッド領域 3 1 の内部にある線分が、エンティティ線分 3 3 として抽出される。図 9 (b) は、抽出されたこのエンティティ線分 3 3 を切り出して (切断して) 候補線分 3 4 を抽出した状態を示している。

【 0 0 2 9 】

図 1 0 は、図 2 のステップ 1 0 7 における有効エッジ候補の選択結果を示した図である。ここでは、図 9 (b) に示した候補線分 3 4 から、有効エッジ候補 3 5 を選択している。この有効エッジ候補 3 5 は、候補線分 3 4 として選ばれたものの中から最終的な領域のエッジとして有効なものが選ばれている。より具体的には、例えば、隣接ピクセル領域 3 2 の単位ピクセルにある候補線分 3 4 の中で、ピクセルグリッド領域 3 1 を構成する単位グリッドエッジから最も距離の近いものを有効エッジ候補 3 5 として選択している。

【 0 0 3 0 】

図 1 1 は、この有効エッジ候補 3 5 の選択手順の流れを示した図である。まず、グリッドエッジを挟む 2 つのピクセル内の候補エッジを選ぶ (ステップ 2 0 1) 。この候補エッジが折れ線である場合は、各頂点の二乗誤差を最小化するような単一線分で近似する (ステップ 2 0 2) 。これにより、候補エッジは全て単一線分となる。次に、グリッドエッジに対する角度が 45° 以内のものを取り出す (ス

テップ 2 0 3)。また、グリッドエッジの線分 $(x_{s1}, y_{s1}) - (x_{t1}, y_{t1})$ と候補エッジの線分 $(x_{s2}, y_{s2}) - (x_{t2}, y_{t2})$ との角度を図に示す式で求める(ステップ 2 0 4)。この式は、値が小さいほど平行に近いことを示している。次に、候補エッジの midpoint とグリッドエッジとの距離を計算する。距離はピクセルのサイズを単位とする数値として求める(ステップ 2 0 5)。そして、角度と距離に適当な重みを乗じ、その数値の小さいものをその単位グリッドに対する有効エッジ候補 3 5 とする(ステップ 2 0 6)。最後に候補の中で近似したものは、元の折れ線に戻す(ステップ 2 0 7)。このように、有効エッジ候補 3 5 としては、グリッドエッジに近いだけでなく、傾きの大きさについても考慮される。即ち、いくらグリッドエッジに近くても傾きが大きければ関係ないと判断され、遠い場合であっても線に沿ったものの場合には有効エッジ候補 3 5 として選択され得る。

【 0 0 3 1 】

図 1 2 は、図 2 のステップ 1 0 8 における有効エッジ候補 3 5 の線分の接続例を示した図であり、符号 3 6 の破線で示される線分が加えられた接続線分である。各単位グリッドに対して有効エッジ候補 3 5 を求めたが、連続する単位グリッドに対する有効エッジが連続していない場合も多い。即ち、連続しない有効エッジ候補 3 5 が選ばれるケースと、対応する有効エッジ候補 3 5 がない場合がある。このような非連続の有効エッジ候補 3 5 に対しては、連続が途絶えた有効エッジ候補 3 5 と次に現れる有効エッジ候補 3 5 を補完関数などによる接続線分 3 6 で滑らかに接続する。

【 0 0 3 2 】

図 1 3 は、図 2 のステップ 1 0 9 におけるひげの除去を説明するための図である。有効エッジ候補 3 5 に対して接続線分 3 6 による接続作業が完了すると、有効エッジ候補 3 5 でポリゴンを作ることができる。しかしながら、このままのポリゴンには、図 1 3 の符号 3 7 で示すような「ひげ」ができています。ひげ 3 7 は、図に示すようなクロス線分の先に伸びる行き止まりになる折れ線である。このようなひげ 3 7 は、クロスする部分を平面走査法などの交点列挙アルゴリズムで求め、交点からの接続関係を調べることで解消することができる。

【 0 0 3 3 】

図 1 4 は、以上のようにして作成された領域を地図エンティティに対応して出力した例である。図 7 のピクセルグリッド領域 3 1 は、実在する地図上のエンティティに合うように調整されて、図 1 4 に示すような調整領域 3 8 に変換され、表示装置 1 6 によって地図に反映されて表示出力される。即ち、抽出されて調整された道路情報等のエンティティや接続線分 3 6 は、例えば、赤色等の色づけや太線、また、調整された閉領域の着色や白黒反転等の強調処理が施され、例えば表示装置 1 6 のディスプレイを用いて、地図上に表示出力される。

このように、本実施の形態では、領域最適化を空間的な連続性を考慮した上で行うことで、より高度な空間データマイニングを提供することが可能となる。また、ピクセルグリッドを用いて計算時間を早めて算出できると共に、その出力では、地図上のエンティティを用いて調整することができ、ユーザにとって使いやすい空間データマイニングを提供することが可能である。例えば、スーパーマーケットが新聞の折り込みチラシを配布する際に、最適な地域を求めるアプリケーションに応用した場合には、川や道路、公園などのエンティティを踏まえて最適な領域を出力することができる。かかる出力を用いて、例えば、各新聞販売店に対してその地域を明らかにして折り込み指示を行えば、効率的な且つ実質価値の高いアウトプットを提供することが可能となる。

【 0 0 3 4 】

【発明の効果】

以上説明したように、本発明によれば、出力の際に要求される領域情報が含まれていない目的関数を用いて、大量のデータベースから領域最適化を図ることができる。また、最適化された領域情報を実在の地図エンティティに合うように調整して、算出された領域情報を地図と共に出力することが可能となる。

【図面の簡単な説明】

【図 1】 本実施の形態における最適領域算出手法を実現するためのコンピュータ・システムの概略構成を説明するための図である。

【図 2】 本実施の形態における最適領域算出方法を実現するための処理の流れを示した図である。

【図 3】 (a),(b)は、HDD 1 5 等に格納されている顧客データベース

例と入力装置 1 7 から入力される制約、目的関数の例を示す図である。

【図 4】 (a),(b)は、図 2 に示したステップ 1 0 1 にて読み込まれる地図とバケット化された地図の例を示す図である。

【図 5】 図 2 に示したステップ 1 0 3 にて、ピクセルグリッド上のバケットに集計されたデータの一例を示す図である。

【図 6】 (a),(b)は、最適化関数値を最も良くする領域の算出例を示す図である。

【図 7】 算出して求められた最適ピクセルグリッド領域を示す図である。

【図 8】 図 2 のステップ 1 0 5 における隣接ピクセル領域 3 2 の計算結果を示す図である。

【図 9】 (a),(b)は、図 2 のステップ 1 0 6 における隣接ピクセルと交差する地図エンティティの切り出し作業を説明するための図である。

【図 1 0】 図 2 のステップ 1 0 7 における有効エッジ候補 3 5 の選択結果を示した図である。

【図 1 1】 有効エッジ候補 3 5 の選択手順の流れを示した図である。

【図 1 2】 図 2 のステップ 1 0 8 における有効エッジ候補 3 5 の線分の接続例を示した図である。

【図 1 3】 図 2 のステップ 1 0 9 におけるひげ 3 7 の除去を説明するための図である。

【図 1 4】 作成された領域を地図エンティティに対応して出力した例を示す図である。

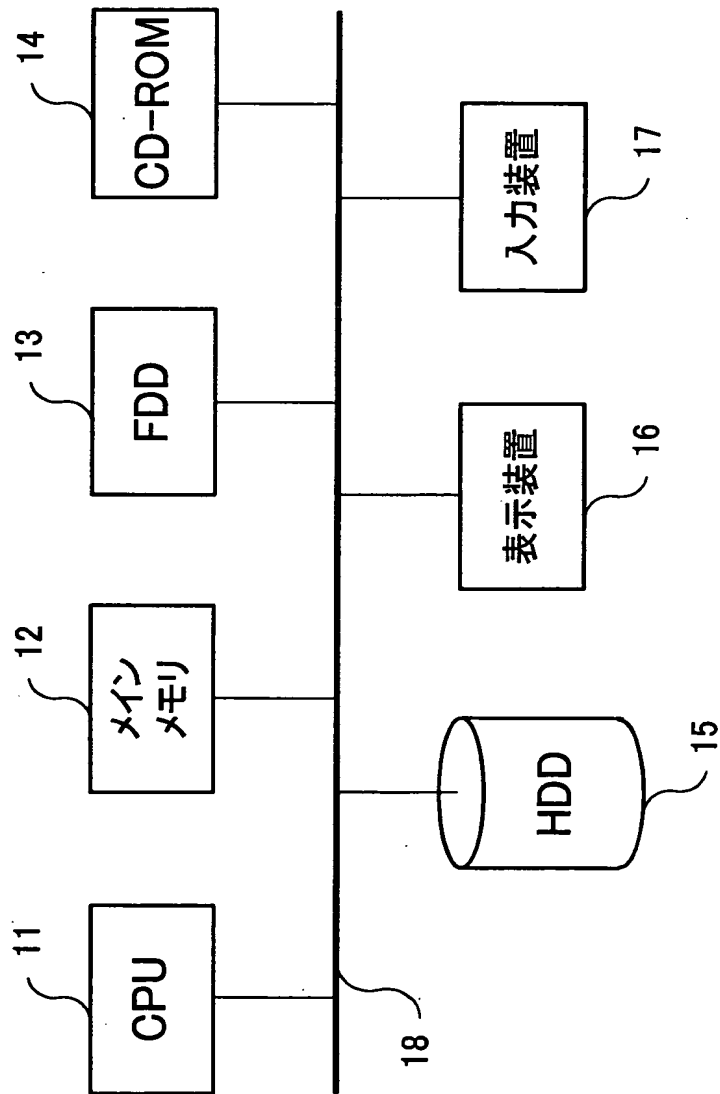
【符号の説明】

1 1 … CPU、1 2 … メインメモリ、1 3 … フロッピーディスクドライブ(FDD)、1 4 … CD-ROMドライブ、1 5 … ハードディスクドライブ(HDD)、1 6 … 表示装置、1 7 … 入力装置、1 8 … バス、3 1 … ピクセルグリッド領域、3 2 … 隣接ピクセル領域、3 3 … エンティティ線分、3 4 … 候補線分、3 5 … 有効エッジ候補、3 6 … 接続線分、3 7 … ひげ、3 8 … 調整領域

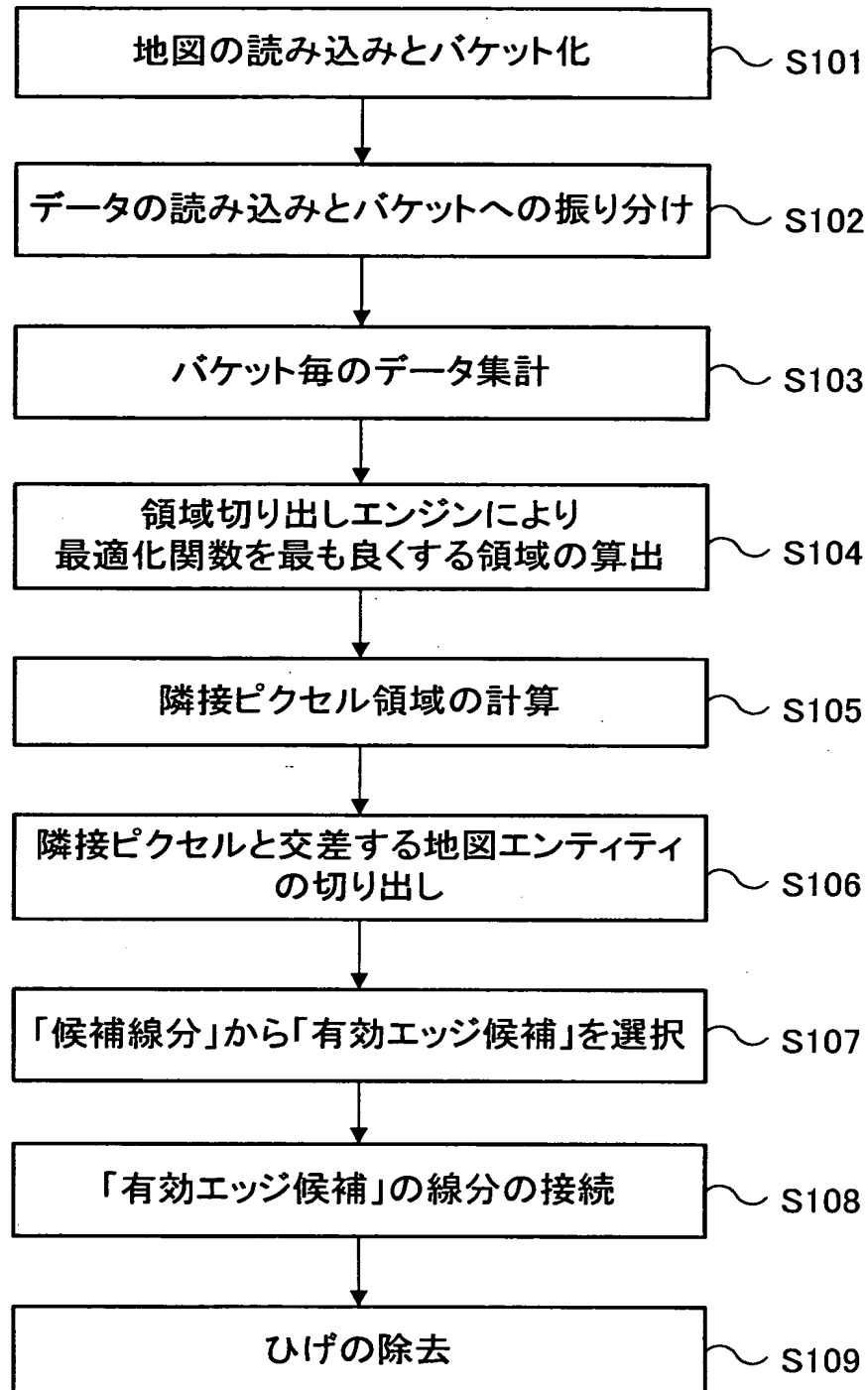
【書類名】

図面

【図 1】



【図 2】



【図 3】

(a) あるデパートの顧客データベース

顧客ID	名前	住所 X , Y	売上金額	年収
101	ichiro	100 , 150	126,800	5,000,000
102	jiro	220 , 180	198,000	7,000,000
103	saburo	420 , 500	15,000	5,800,000
⋮	⋮	⋮	⋮	⋮

(b) 領域最適化のための制約、目的関数

制約 :

目的関数 :

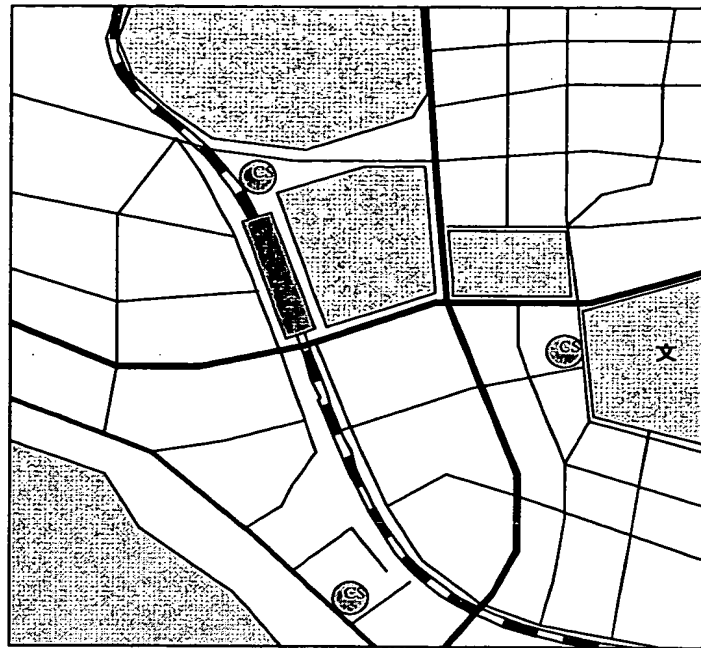
最大グリット数

売上金額の総計

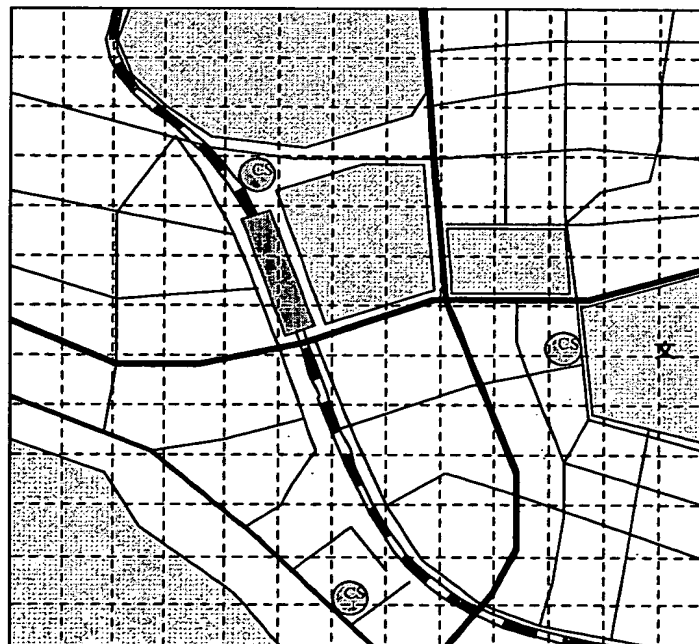
最小顧客数

平均年収

【図4】

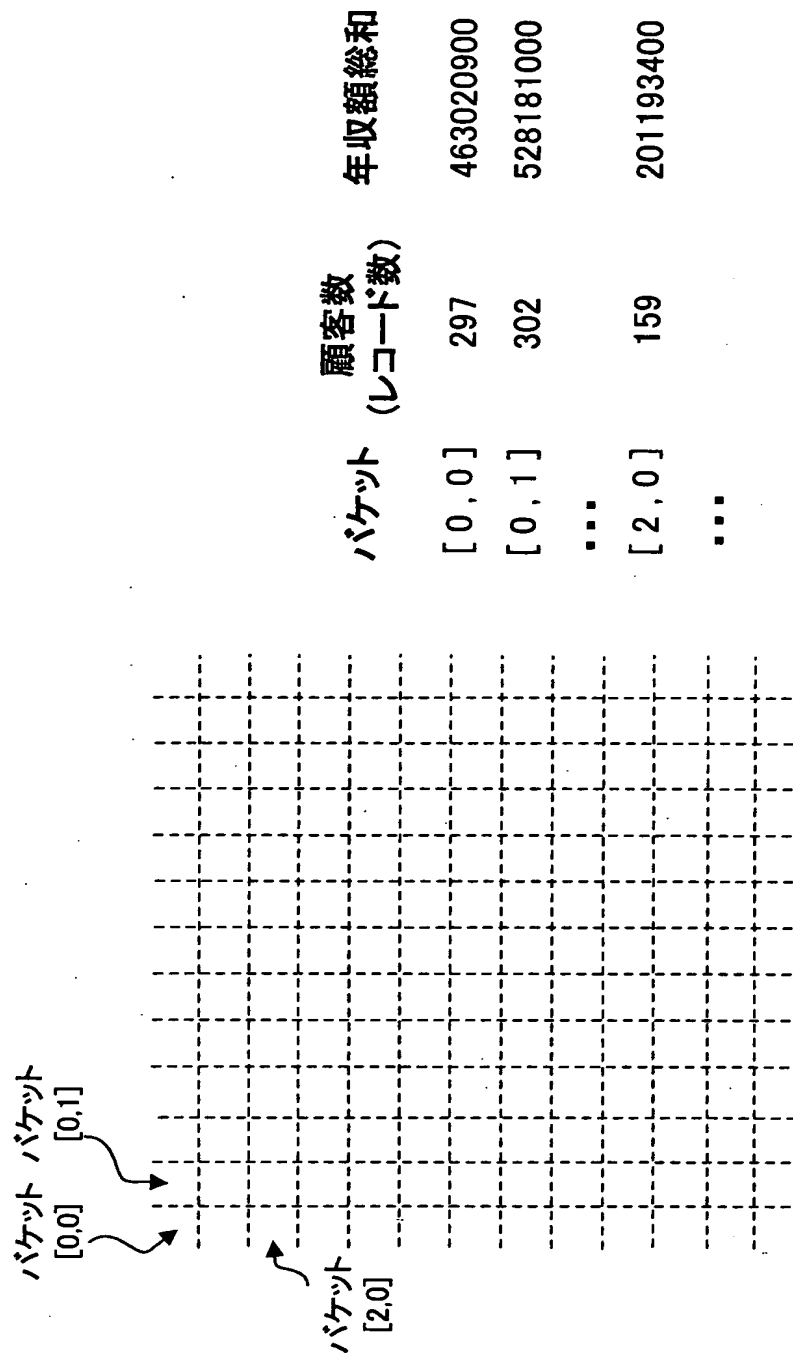


(a) 地図

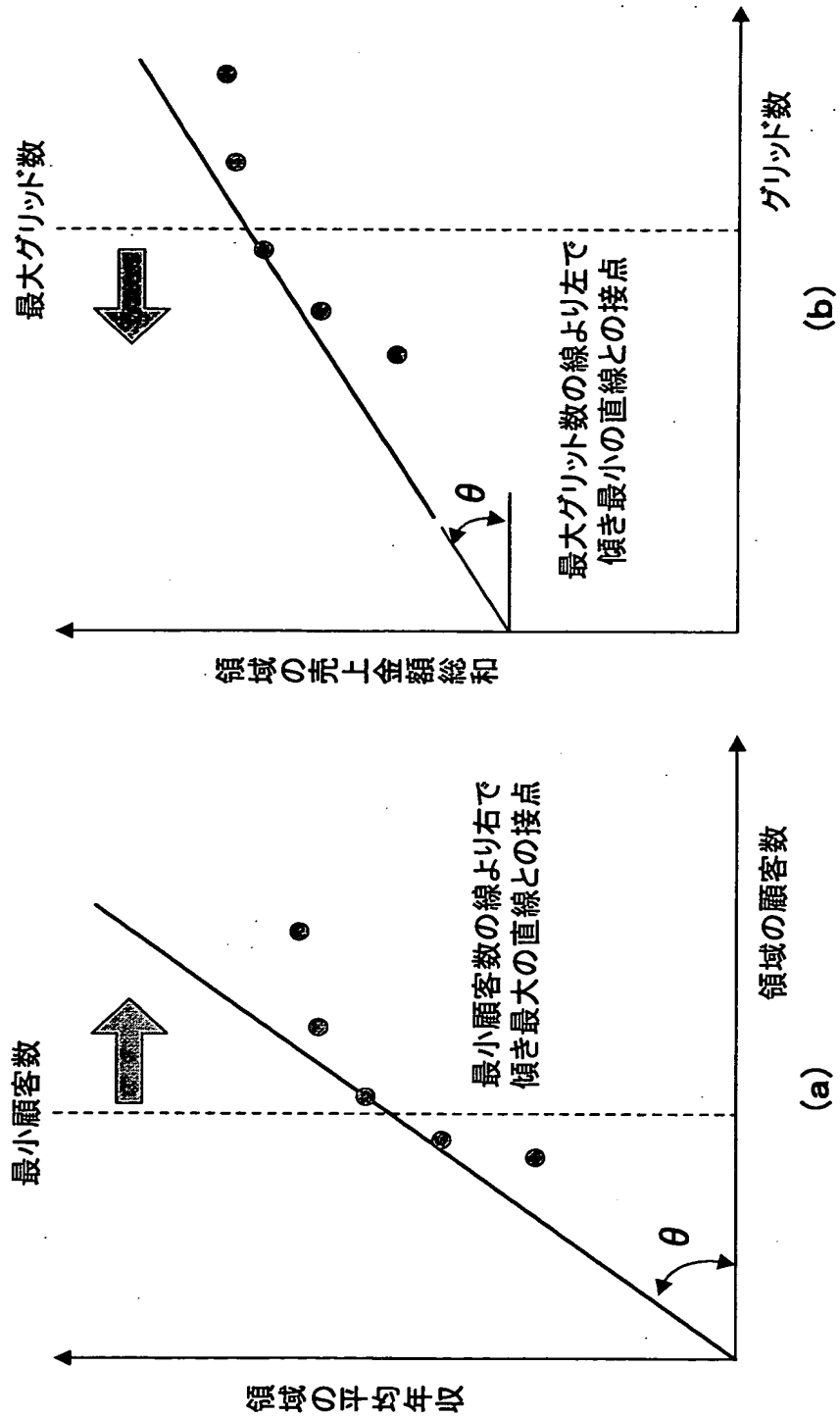


(b) バケット化された地図

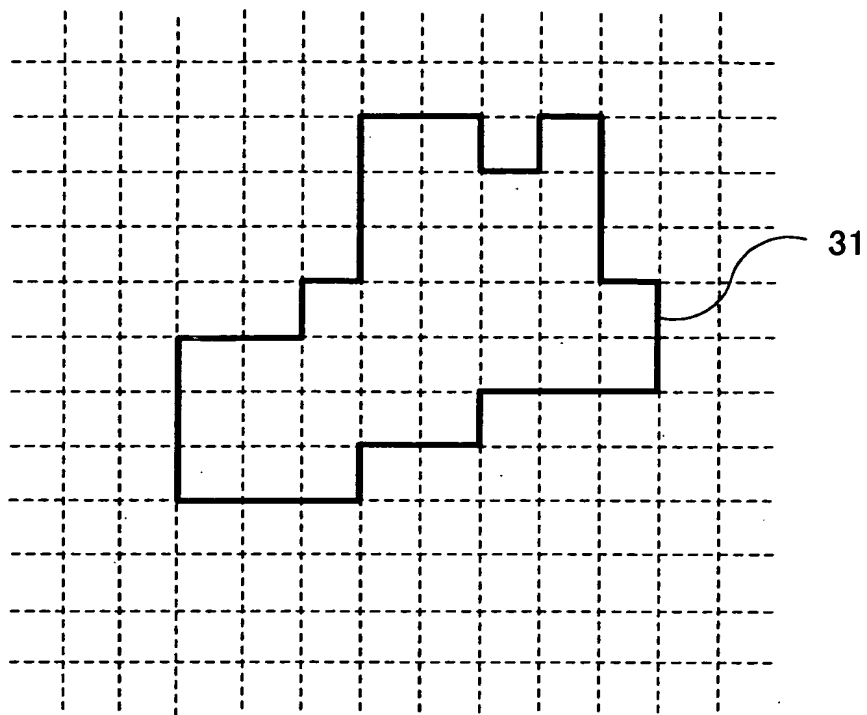
【図 5】



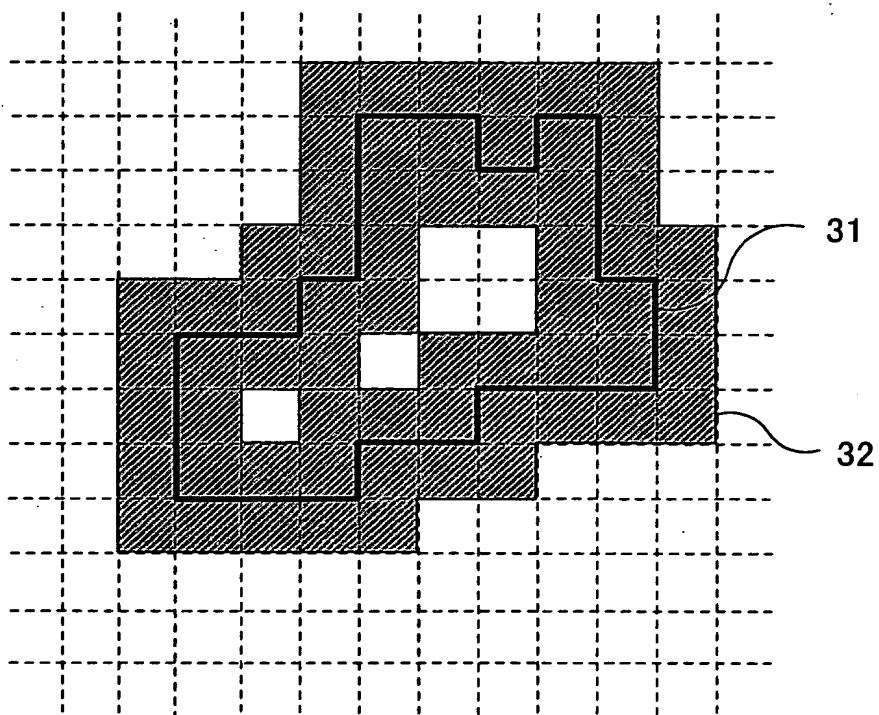
【図 6】



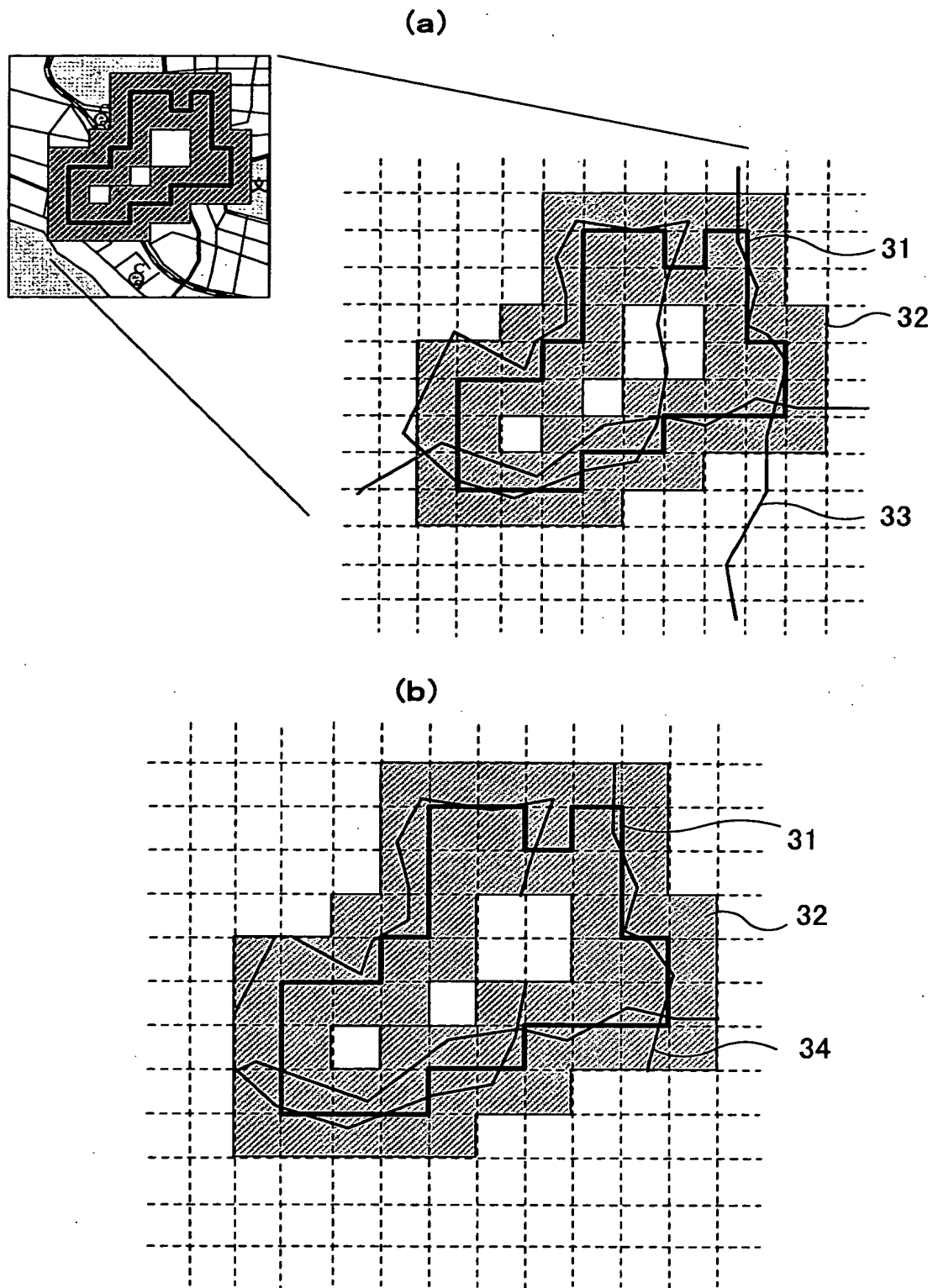
【図 7】



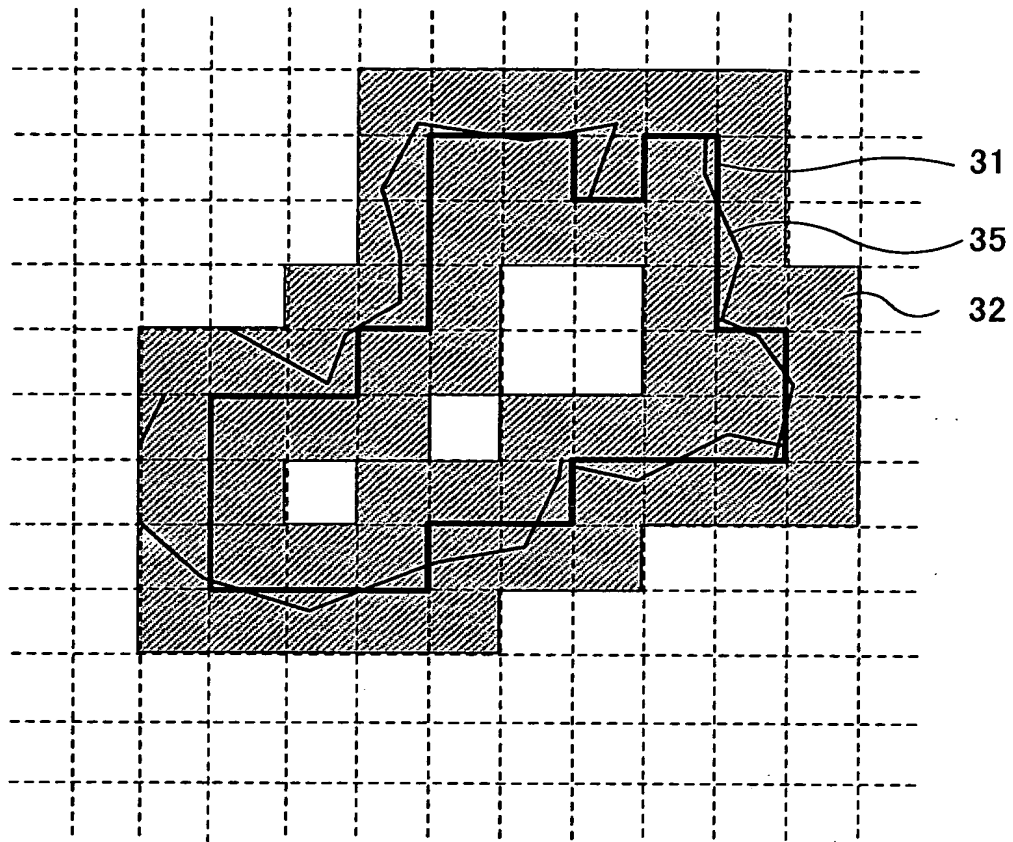
【図 8】



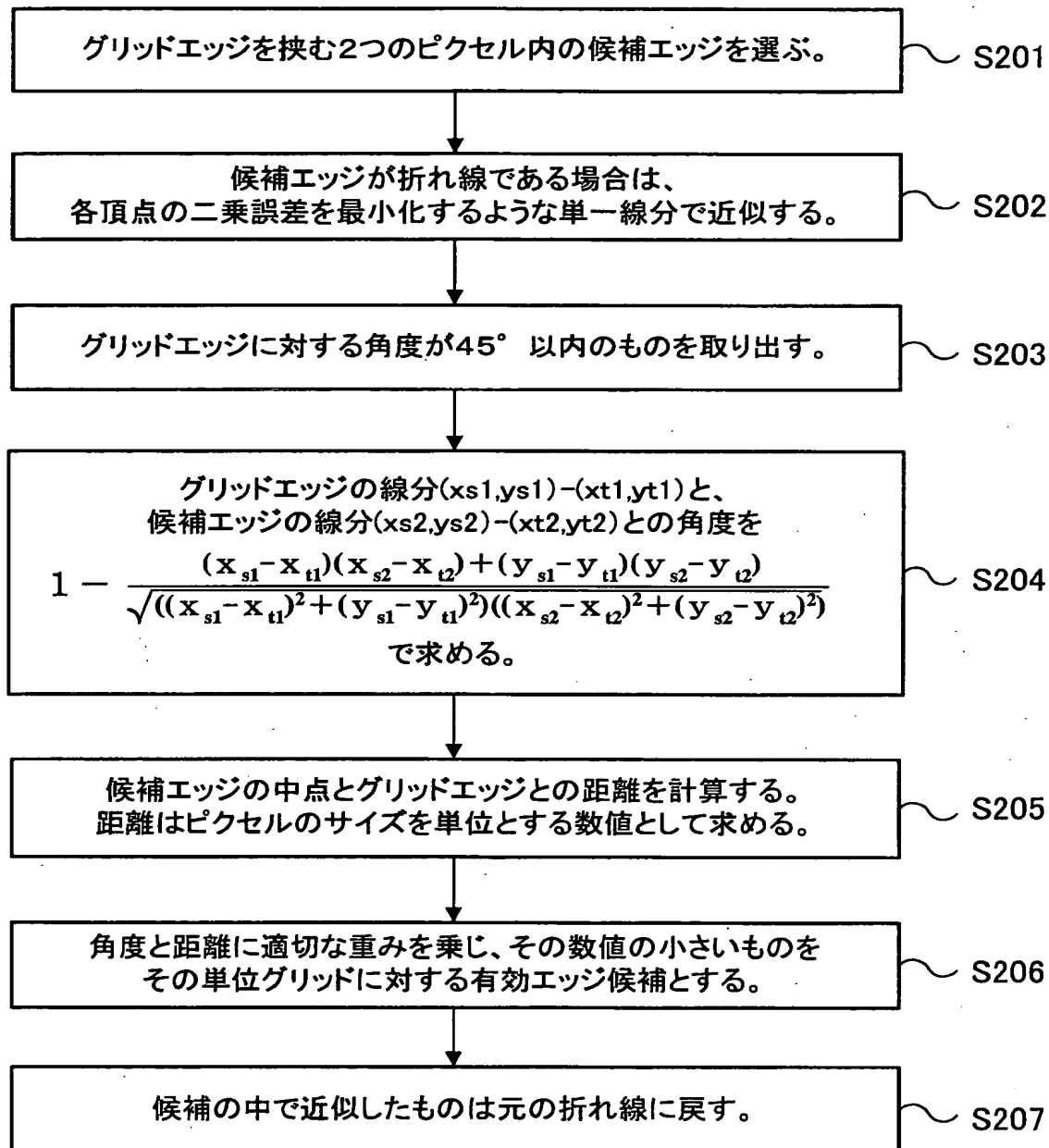
【図 9】



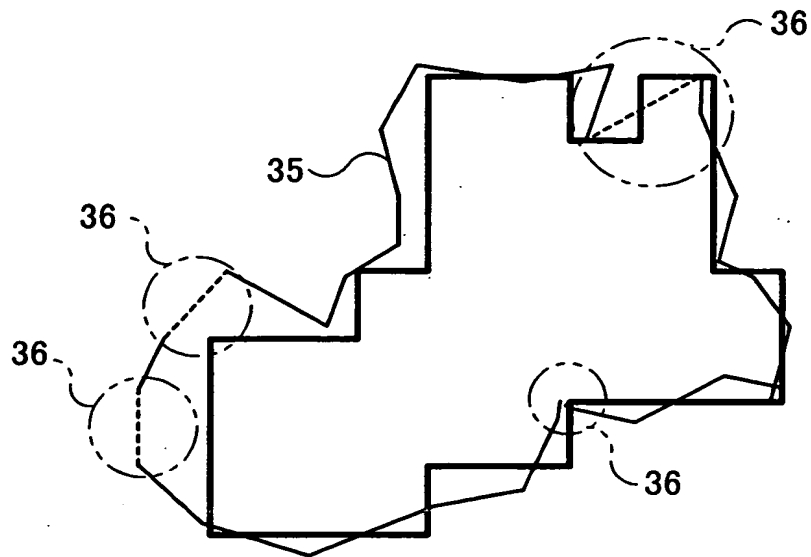
【図 1 0】



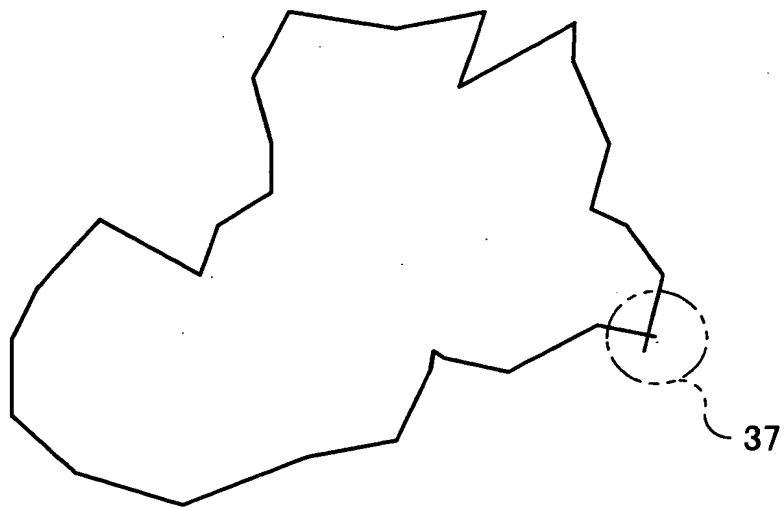
【図 1 1】



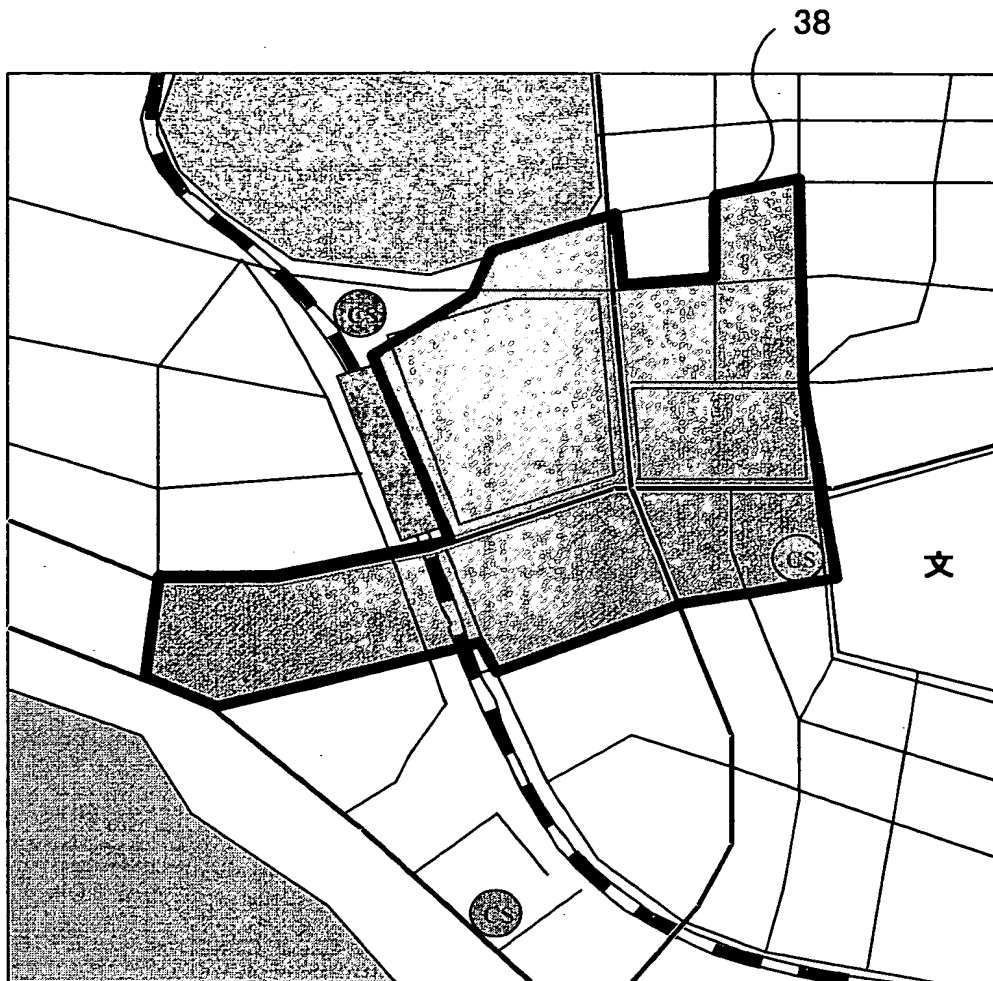
【図 1 2】



【図 1 3】



【図 14】



【書類名】 要約書

【要約】

【課題】 空間的な連続性を考慮した上で領域最適化を図り、より高度な空間データマイニングを提供する。

【解決手段】 住所などの空間情報を含むデータベースの中から2次元相関ルールを導き出して地図上で適用する領域算出方法であって、2次元相関ルールを導き出すために調べる、出力の際に要求される領域情報が含まれていない目的関数を定義するステップと、地図上の領域を所定の大きさからなるピクセルグリッドにバケット化するステップ(S 1 0 1)と、バケット毎にデータベースからデータを集計するステップ(S 1 0 3)と、定義された目的関数に基づいてその目的関数を最適化する領域を算出するステップ(S 1 0 4)と、算出された領域に対応する地図上にあるエンティティを抽出するステップ(S 1 0 6)と、抽出されたエンティティに基づいて地図上で適用する領域を出力するステップとを備える。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願 2 0 0 0 - 1 5 3 0 8 7
受付番号	5 0 0 0 0 6 3 9 4 0 4
書類名	特許願
担当官	濱谷 よし子 1 6 1 4
作成日	平成 1 2 年 7 月 3 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国 1 0 5 0 4、ニューヨーク州 アーモンク (番地なし)
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【復代理人】

【識別番号】	申請人 100104880
【住所又は居所】	東京都港区赤坂 7 - 1 0 - 9 第 4 文成ビル 2 0 2 セリオ国際特許事務所
【氏名又は名称】	古部 次郎

【選任した代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

【選任した復代理人】

【識別番号】	100100077
【住所又は居所】	東京都港区赤坂 7 - 1 0 - 9 第 4 文成ビル 2 0 2 セリオ国際特許事務所
【氏名又は名称】	大場 充

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 2000年 5月16日
[変更理由] 名称変更
住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション